

# Ab initio analysis of all income society classes in the European Union

M. JAGIELSKI<sup>†</sup>, R. KUTNER

Institute of Experimental Physics  
Faculty of Physics, University of Warsaw  
Hoża 69, PL-00681 Warsaw, Poland

We found a unified formula for description of the household incomes of all society classes, for instance, of those of the European Union in year 2007. This formula is a stationary solution of the threshold Fokker-Planck equation (derived from the threshold nonlinear Langevin one). The formula is more general than the well known that of Yakovenko et al. because it satisfactorily describes not only household incomes of low- and medium-income society classes but also the household incomes of the high-income society class.

PACS numbers: 89.20.-a, 89.65 Gh

Final draft submitted to Acta Physica Polonica A

## 1. Introduction

In study of socio-economical systems, physics oriented approaches have widely been developed to explain different socio-economic processes [1–8].

---

<sup>†</sup> e-mail: zagielski@interia.pl

Those approaches aim at formulating well fitted unbiased indicators of social and economic phenomena. One of their key issues is the income of society analysis using methods of statistical physics, in particular, the stochastic dynamics considered as *ab initio* level. The main goal of this economic issue is to unravel and describe mechanisms of societies' enrichment or impoverishment.

In the recent decade, a large number of studies were performed aiming at constructing of models, which (to some extend) would well replicate the observed complementary cumulative distribution functions of individual incomes. Among them, the most significant seems to be the Clementi-Matteo-Gallegati-Kaniadakis approach [9], the Generalized Lotka-Volterra Model [4–6], the Boltzmann-Gibbs law [10–13], and the Yakovenko et al. model [2, 3]. However, none of the above attempts to find an analytical description of the income structure solves the principal challenges, which concern:

- (i) the description of the annual household incomes of all society classes (including the third, i.e. the high-income society class) by a single unified formula based on the *ab initio* level and
- (ii) the problem regarding corresponding complete microscopic (microeconomic) mechanism responsible for the income structure and dynamics.

In our considerations presented herein, we used Boltzmann-Gibbs law, weak Pareto law and Yakovenko et al. model to derive a uniform analytical formula describing all three society classes.

## 2. Extended Yakovenko et al. model

In accord with an effort outlined above, we compared the empirical data of the annual household incomes in the European Union (EU), including Norway and Iceland, with predictions of our theoretical approach proposed herein. This approach is directly inspired by the Yakovenko et al. model. By using the generalised assumptions we extended this model to solve our principal challenges (i) and (ii) indicated above.

We used data records from the Eurostat Survey on Income and Living Conditions (EU-SILC) [14], by way of example for year 2007 [15] (containing around 200 thousand empirical data points). However, these records contain only few data points concerning the high-income society class, i.e. the third region in the plot of the complementary cumulative probability distribution function vs. annual household income. To consider the high-income society class systematically, we additionally analysed the effective income of billionaires<sup>1,2</sup> in the EU by using the Forbes 'The World's Billionaires' rank [16].

We were able to consider incomes of three society classes thanks to the following procedure.

(i) Firstly, we selected EU billionaires' wealth from the Forbes rank, for instance, for two successive years 2006 and 2007.

(ii) Secondly, we calculated their incomes for year 2007. This calcula-

---

<sup>1</sup> The term 'billionaire' used herein is equivalent (as in the US terminology) to the term 'multimillionaire' used in the European terminology. Since we consider wealth and income of billionaires in euros, we recalculated US dollars to euros by using the mean exchange rate at the day of construction of the Forbes 'The World's Billionaires'.

<sup>2</sup> The billionaires who gained effective incomes are billionaires whose incomes are greater than zero.

tion was possible because we assumed that billionaires' incomes were proportional to the corresponding differences between their wealth for pair of successive years, here 2007 and 2006. Notably, we took into account only billionaires who gained effective incomes.

- (iii) Subsequently, having calculated incomes for the high-income society class, we joined them with the EU-SILC dataset. By using so completed dataset, we then constructed the initial empirical complementary cumulative distribution function for year 2007. For that, we used the well known Weibull recipe [17, 18]. However, this direct approach shows a wide gap of incomes inside the high-income society class resulting in a horizontal line of the complementary cumulative distribution function. This gap separates the first segment belonging to the high-income society class, consisting of all data points taken from the EU-SILC dataset, from the second segment, consisting of remaining data points, which also belong to the high-income society class but are taken from the Forbes dataset.
- (iv) In the final step, we eliminated this gap by adopting the assumption that the empirical complementary cumulative distribution function (concerning the whole society) have no horizontal segments. That is, we assumed that statistics of incomes is a continuous function of income (i.e. it has no disruption). Hence, we were forced to multiply the billionaire incomes from Forbes dataset by the properly chosen common proportionality factor. This factor was equal to  $1.0 \times 10^{-2}$ , as we assumed the requirement of full overlap of the first (above mentioned) segment by the second segment. This assumption leads to a unique solution (up to some negligible statistical error) for this proportionality

factor. We found that this factor was only a slowly-varying function of time (or years).

Hence, we received data record containing already a sufficient number of data points for all society classes, including the high-income society class. Although the Forbes empirical data only roughly estimate the wealth of billionaires, they quite well establish the billionaires' rank, thus sufficiently justifying our approach. This is because our purpose is to classify billionaires to concrete universality class rather than finding their total incomes.

The basic tool of our analysis is an empirical complementary cumulative distribution function being typical in this context. We calculated it according to the standard two-step procedure based on the well known Weibull formula [17, 18]. The complementary cumulative distribution function obtained that way is sufficiently stable and it does not reduce the size of the output compared to that of the original empirical data record.

Let  $m$  be an influx of income per unit time to a given household. We treat  $m$  as a variable obeying stochastic dynamics. Then, we can describe its time evolution by using the nonlinear Langevin stochastic dynamics equation [2, 3, 19]. Hence, this Langevin equation is equivalent to the following Fokker-Planck equation for the probability distribution function (in the Itô representation) [19]:

$$\frac{\partial}{\partial t}P(m, t) = \frac{\partial}{\partial m}[A(m)P(m, t)] + \frac{\partial^2}{\partial m^2}[B(m)P(m, t)]. \quad (1)$$

Here,  $A(m)$  is a drift coefficient and  $B(m) = C^2(m)/2$ , where the coefficient  $C(m)$  is the  $m$ -dependent amplitude of a temporal white noise; they together play a fundamental role in the Langevin equation as a stochastic force. The quantity  $P(m, t)$  is the temporal income distribution function. In general, coefficients  $A(m)$  and  $B(m)$  can be additionally determined by the first and

second moment of the income change per unit time, respectively, only if these moments exist. Subsequently, the equilibrium solution of Eq. (1),  $P_{\text{eq}}$ , takes the form:

$$P_{\text{eq}}(m) = \frac{\text{const}}{B(m)} \exp \left( - \int_{m_{\text{init}}}^m \frac{A(m')}{B(m')} dm' \right) \quad (2)$$

where  $m_{\text{init}}$  is the lowest household income and *const* is a normalisation factor. Indeed, this expression is exploited in this work.

Following the Yakovenko et al. model [2,3], we can assume that changes of income of the low-income society class are independent of the previous income gained. This assumption is justified because the income of households belonging to this class mainly takes the form of wages and salaries. The stochastic process associated with the mechanism of this kind is called the additive stochastic process. In this case, coefficients  $A(m)$  and  $B(m)$  take, obviously, the form of positive constants

$$A(m) = A_0, \quad B(m) = B_0. \quad (3)$$

This choice of coefficients leads to the Boltzmann-Gibbs law with exponential complementary cumulative distribution function [2,3,10–13]:

$$\Pi(m) = \int_m^\infty P_{\text{eq}}(m') dm' = \exp \left( - \frac{m - m_{\text{init}}}{T} \right). \quad (4)$$

In Equation (4), distribution function is characterised by a single parameter, i.e. an income temperature  $T = B_0/A_0$ , which can be interpreted in this case as an average income per household.

For the medium- and high-income society classes, we can assume (again following Yakovenko et al. [2,3]) that changes of income are proportional to the income gained so far. This assumption is also justified because profits go to the medium- and high-income society classes mainly through investments

and capital gains. This type of stochastic process is called the multiplicative stochastic process. Hence, coefficients  $A(m)$  and  $B(m)$  obey the proportionality principle of Gibrat [20, 21]:

$$A(m) = a m, \quad B(m) = b m^2 \Leftrightarrow C(m) = \sqrt{2b} m, \quad (5)$$

where  $a$  and  $b$  are positive parameters. By using the equilibrium distribution function, Eq. (2), we arrive in this case to the weak Pareto law with complementary cumulative distribution function [2, 3, 6]:

$$\Pi(m) = \int_m^\infty P_{\text{eq}}(m') dm' = \left( \frac{m}{m_s} \right)^{-\alpha}. \quad (6)$$

Here,  $m_s$  is a scaling factor (depending on  $a$ ,  $b$ , and *const*) while  $\alpha = 1 + a/b$  is the Pareto exponent. The ratio of the  $a$  to  $b$  parameters can directly be determined from the empirical data expressed in the log-log plot (by using their slopes).

As Yakovenko et al. have already found [2, 3], the coexistence of additive and multiplicative stochastic processes is allowed. By assuming that these processes are uncorrelated, we get

$$A(m) = A_0 + a m, \quad B(m) = B_0 + b m^2 = b (m_0^2 + m^2), \quad (7)$$

where  $m_0^2 = B_0/b$ . This consideration leads (together with Eq. (2)) to a significant Yakovenko et al. model with the probability distribution function given by

$$P_{\text{eq}}(m) = \text{const} \frac{e^{-(m_0/T) \arctan(m/m_0)}}{[1 + (m/m_0)^2]^{(\alpha+1)/2}}, \quad (8)$$

where parameters  $\alpha$  and  $T$  are defined above.

Based on the Yakovenko et al. Eq. (8), the complementary cumulative distribution function can describe income of only low- and medium-income

society classes. However, it does not capture that of the most intriguing high-income society class.

The goal of our present work is to derive from Eq. (2) such a distribution function, which would cover all three ranges of the empirical data records, i.e. low-, medium, and high-income classes of the society (including also two short intermediate regions between them). To do that, we have to provide function  $A(m)$  in the threshold form:

$$\begin{aligned} A(m) &= \begin{cases} A^<(m) = A_0 + a m & \text{if } m < m_1 \\ A^>(m) = A'_0 + a' m & \text{if } m \geq m_1, \end{cases} \\ B(m) &= B_0 + b m^2 = b(m_0^2 + m^2). \end{aligned} \quad (9)$$

At the threshold  $m_1$ , there is a jump of the proportionality coefficient of the drift term. That is, this term abruptly changes from  $a$  to  $a'$  while the formalism of the income change remains the same for the whole society. This formalism is expressed by the threshold nonlinear Langevin equation where particular dynamics distinguishes the range of the high-income society class from those of the others.

The threshold parameter  $m_1$  can be interpreted as a crossover income between the medium- and high-income society classes. Remarkably, both income crossovers  $m_0$  and  $m_1 (\geq m_0)$  are exogenous parameters. They should be determined from the dependence of the empirical complementary cumulative distribution function on variable  $m$  because both crossovers are sufficiently distinct.

Subsequently, by substituting Eq. (9) into Eq. (2), we finally get

$$P_{\text{eq}}(m) = \begin{cases} c' \frac{\exp(-(m_0/T) \arctan(m/m_0))}{[1+(m/m_0)^2]^{(\alpha+1)/2}}, & \text{if } m < m_1 \\ c'' \frac{\exp(-(m_0/T_1) \arctan(m/m_0))}{[1+(m/m_0)^2]^{(\alpha_1+1)/2}}, & \text{if } m \geq m_1 \end{cases} \quad (10)$$



where  $\alpha_1 = 1 + a'/b$  and  $T_1 = B_0/A'_0$ . Apparently, the number of free (effective) parameters driving the two-branch distribution function, Eq.(10), is reduced because this function depends only on the ratio of the initial parameters defining the nonlinear Langevin dynamics.

For  $m_1 \gg m_0$ , the interpretation of the distribution function, Eq. (10), is self-consistent, as required, because the two power-law regimes are well defined. Then, for instance for  $m \gg m_0$ , the second branch in Eq. (10) becomes the power-law dependence driven by the Pareto exponent  $\alpha_1$  different (in general) from  $\alpha$ .

Importantly, our analysis indicates that the existence of the third income region is already allowed by theory. We are following this indication below.

### 3. Results and discussion

In principle, we are ready to compare the theoretical complementary cumulative distribution function based on our probability distribution function  $P_{\text{eq}}(m)$ , given by Eq. (10), with the empirical data for the whole income range. However, the analytical form of this theoretical complementary cumulative distribution function is unknown in the closed explicit form. Therefore, we calculate it numerically. The key technical question arises on how to fit this complicated theoretical function to the empirical data. The fitting procedure consists of three steps as, fortunately, all parameters are to be found (in principle) by using independent fitting routines, as follows.

In an initial step, we found approximated values of crossovers  $m_0$  and  $m_1$  directly from the plot of the empirical complementary cumulative distribution function (or empirical data). Thus, uncertainty of the  $m_0$  and  $m_1$  parameters did not exceed 10%, which was sufficiently accurate. Moreover,

we took the exact value of the parameter  $m_{\text{init}}$  as the first point in the record of the empirical data.

Secondly, we determined the temperature  $T$  value by fitting the Boltzmann-Gibbs formula, Eq. (4), to the corresponding empirical data in the range extending from  $m_{\text{init}}$  to  $m_0$  (both found in the initial step). Notably, we assumed that this formula could be characterised by a single temperature value since the society as a whole was considered to be in (partial) equilibrium during the whole fiscal year. That is, we further put  $T_1 = T \Leftrightarrow A'_0 = A_0$ .

At the third step, we determined exponents  $\alpha$  and  $\alpha_1$  by separately fitting the weak Pareto law to the empirical data for the medium- and high-income society classes, respectively.

Hence, we have already obtained all values required by the extended Yakovenko et al. formula, Eq. (10). The corresponding plots of the empirical and theoretical complementary cumulative distribution functions in the log-log scale are compared in Fig. 1, for instance, for year 2007. Apparently, the predictions of the extended Yakovenko et al. formula, Eq. (10), (solid curve in Fig. 1) well agree with the empirical data (dots in Fig. 1) for low- and medium-income society classes while agreement for the high-income society class is satisfactory.

#### 4. Concluding remarks

Herein, we proved that the household incomes of all society classes in the EU can be modelled by the nonlinear threshold Langevin dynamics with  $m$ -dependent drift and dispersion as *ab initio* level. At the threshold  $m_1$ , there is a jump of the proportionality coefficient of the drift term. That is, this term abruptly changes from  $a$  to  $a'$ , where  $a' < a$  (as  $\alpha_1 < \alpha$ ). It

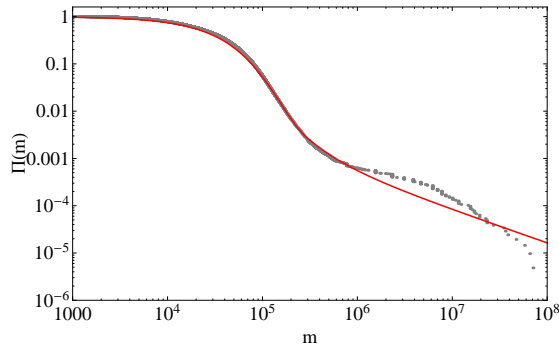


Fig. 1. Fit of the complementary cumulative distribution function, based on the extended Yakovenko et al. formula, Eq. (10), (solid line) to the EU household income empirical data set (dots) for year 2007 ( $T_1 = T_2 = T = 37 \times 10^3 \pm 1 \times 10^3$  EUR,  $m_0 = 1.60 \times 10^5 \pm 0.16 \times 10^5$  EUR,  $m_1 = 3 \times 10^5 \pm 0.3 \times 10^5$  EUR,  $\alpha = 2.8643 \pm 0.0008$ , and  $\alpha_1 = 0.70 \pm 0.02$ ) [15, 16].

means that the stochastic term in the Langevin equation is relatively more significant in this case (i.e. above threshold  $m_1$ ) than the drift term.

Furthermore, for the medium-income society class the Pareto exponent  $\alpha > 2$ . This means that the variance of the Pareto distribution function exists and it is finite. However, for the high-income society class the variance of the Pareto distribution function is infinite, because  $\alpha_1 < 1$ . That is, assuming the variance as a measure of a risk, the economic activity of the high-income society class can be considered as more risky than activities of all other society classes, as expected [1].

The completed database, which we used (by properly joining the Forbes empirical database with that of EU-SILC), emphasises a significant role of the high-income society class. That is, only study of the income of all society classes enables adequate characterisation of the relative society wealth.

## Acknowledgements

We thank Victor M. Yakovenko and Tiziana Di Matteo for very stimulating comments and suggestions.

## REFERENCES

- [1] R. N. Mantegna, H.E. Stanley, *An Introduction to Econophysics: Correlations and Complexity in Finance*, Cambridge University Press, Cambridge, 2000.
- [2] V. M. Yakovenko, J. B. Rosser, *Rev. Mod. Phys.* **81**, 1703 (2009).
- [3] A. Banerjee, V. M. Yakovenko, *New J. Phys.* **12**, 075032 (2010).
- [4] S. Solomon, P. Richmond, *Eur. Phys. J. B* **27**, 257 (2002).
- [5] P. Richmond, S. Solomon, *Int. J. Mod. Phys. C* **12**, 333 (2001).
- [6] P. Richmond, S. Hutzler, R. Coelho, P. Repetowicz, in: *Econophysics and Sociophysics: Trends and Perspectives*, Eds. B. K. Chakrabarti, A. Chakraborti, A. Chatterjee, WILEY-VCH, Weinheim, 2006, p. 131.
- [7] M. Jagielski, R. Kutner, *Acta Phys. Pol. A* **117**, 615 (2010).
- [8] M. Jagielski, R. Kutner, *Acta Phys. Pol. A* **121**, B-47 (2012).
- [9] F. Clementi, T. Di Matteo, M. Gallegati, G. Kaniadakis, *Physica A* **387**, 3201 (2008).
- [10] A. Chatterjee, B. K. Chakrabarti, *Eur. Phys. J. B* **60**, 135 (2007).
- [11] P. Bhattacharyya, A. Chatterjee, B. K. Chakrabarti, *Physica A* **381**, 377 (2005).
- [12] A. Dragulescu, V. M. Yakovenko, *Eur. Phys. J. B* **17**, 723 (2000).
- [13] A. Dragulescu, V. M. Yakovenko, *Eur. Phys. J. B* **20**, 585 (2001).
- [14] [http://epp.eurostat.ec.europa.eu/portal/page/portal/microdata/eu\\_silc](http://epp.eurostat.ec.europa.eu/portal/page/portal/microdata/eu_silc)
- [15] EUSILC UDB 2007 – version 3 of March 2010.

- [16] The World's Billionaires: <http://www.forbes.com/wealth/billionaires>
- [17] W. C. Haneberg, *Computational geosciences with Mathematica*, part 1, Springer, Berlin, 2004.
- [18] V. Chow et al., *Applied Hydrology*, McGraw-Hill, Sydney, 1988.
- [19] N.G. Van Kampen, *Stochastic Processes in Physics and Chemistry*, Elsevier, Amsterdam, 2007.
- [20] M. Armatte, *Mathématiques et sciences humaines* **129**, 5 (1995).
- [21] J. Sutton, *J. Econ. Lit.* **35**, 40 (1997).